PORTAL

USPTO

Subscribe (Full Service) Register (Limited Service, Free) Login

Search: ● The ACM Digital Library ○ The Guide

fitness unfit threshold predetermined value evaluation docume SEARCH

THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction

Terms used
fitness unfit threshold predetermined value evaluation documents character strings retrieving smilar docu

Sort results by relevance

Display results expanded form

Save results to a Binder

Search Tips

□ Open results in a new window

Try an Advanced Search
Try this search in The ACM G

Results 21 - 40 of 200 Result page: previous 1 **2** 3 4 5 6 7 8 9 10 next

Best 200 shown

Relevance s

21 Sources of Success for Boosted Wrapper Induction
David Kauchak, Joseph Smarr, Charles Elkan
December 2004 **The Journal of Machine Learning Research**, Volume 5
Publisher: MIT Press
Full text available: pdf(281.46 KB) Additional Information: full citation, abstract, index terms

In this paper, we examine an important recent rule-based information extraction (IE) technique
Boosted Wrapper Induction (BWI) by conducting experiments on a wider variety of tasks than p
studied, including tasks using several collections of natural text documents. We investigate
systematically how each algorithmic component of BWI, in particular boosting, contributes to its
We show that the benefit of boosting arises from the ability to reweight examples to learn speci

22 Information access in the presence of OCR errors
Kazem Taghva, Thomas Nartker, Julie Borsack
November 2004 **Proceedings of the 1st ACM workshop on Hardcopy document processing**
Publisher: ACM Press
Full text available: pdf(139.50 KB) Additional Information: full citation, abstract, references, index terms

Over the last 15 years, the Information Science Research Institute (ISRI) at the University of N
Las Vegas (UNLV) has conducted information access research in the presence of OCR errors. Ou
research has focused on issues associated with the construction of large document databases. I
paper, we will highlight our findings and detail our current activities.

Keywords: categorization, document conversion, information extraction, markup

23 NLP: Web-based acquisition of Japanese katakana variants
Takeshi Masuyama, Hiroshi Nakagawa
August 2005 **Proceedings of the 28th annual international ACM SIGIR conference on Resea development in information retrieval SIGIR '05**
Publisher: ACM Press
Full text available: pdf(313.65 KB) Additional Information: full citation, abstract, references, index terms

This paper describes a method of detecting Japanese Katakana variants from a large corpus. Ka
words, which are mainly used as loanwords, cause problems with information retrieval and so o

PORTAL

USPTO

Subscribe (Full Service) Register (Limited Service, Free) Login

Search: ● The ACM Digital Library ○ The Guide

predetermined value evaluation documents character strings r SEARCH

THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

Terms used **predetermined value evaluation documents character strings retrieving smilar documents**

Found **45,517** of **171,143**

Sort results by: relevance

Display results: expanded form

Save results to a Binder

Search Tips

□ Open results in a new window

Try an Advanced Search

Try this search in The ACM Guide

Results 1 - 20 of 200

Best 200 shown

Result page: **1** 2 3 4 5 6 7 8 9 10 next

Relevance scale □ ▁ ▃ ▅ ■

1 Technique for automatically correcting words in text

Karen Kukich

December 1992 **ACM Computing Surveys (CSUR)**, Volume 24 Issue 4

Publisher: ACM Press

Full text available: pdf(6.23 MB)

Additional Information: full citation, abstract, references, citings, index terms, review

Research aimed at correcting words in text has focused on three progressively more difficult problems:(1) nonword error detection; (2) isolated-word error correction; and (3) context-dependent work correction. In response to the first problem, efficient pattern-matching and n-gram analysis techniques have been developed for detecting strings that do not appear in a given word list. In response to the second problem, a variety of general and application-specific spelling cor ...

Keywords: n-gram analysis, Optical Character Recognition (OCR), context-dependent spelling correction, grammar checking, natural-language-processing models, neural net classifiers, spell checking, spelling error detection, spelling error patterns, statistical-language models, word recognition and correction

2 From text to hypertext by indexing

Airi Salminen, Jean Tague-Sutcliffe, Charles McClellan

January 1995 **ACM Transactions on Information Systems (TOIS)**, Volume 13 Issue 1

Publisher: ACM Press

Full text available: pdf(1.98 MB)

Additional Information: full citation, abstract, references, citings, index terms, review

A model is presented for converting a collection of documents to hypertext by means of indexing. The documents are assumed to be semistructured, i.e., their text is a hierarchy of parts, and some of the parts consist of natural language. The model is intended as a framework for specifying hypertextual reading capabilities for specific application areas and for developing new automated tools for the conversion of semistructured text to hypertext. In the model, two well-known paradigms— ...

Keywords: constrained grammars, grammars, hypertext, properties, structured text, test types, text entities, transient hypergraphs

3 Chinese information retrieval based on terms and relevant terms

Yang Lingpeng, Ji Donghong, Tang Li, Niu Zhengyu

September 2005 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 4 Issue 3

Publisher: ACM Press

Full text available: pdf(316.86 KB) Additional Information: full citation, abstract, references, index terms

In this article we describe our approach to Chinese information retrieval, where a query is a short natural language description. First, we use automatically extracted short terms from document sets to build indexes and use the short terms in both the query and documents to do initial retrieval. Next, we use long terms extracted from the document collection to reorder the top *N* retrieved documents to improve precision. Finally, we acquire the relevant terms of the short terms from the Int ...

Keywords: Term extraction, document re-ranking, information retrieval, query expansion, relevant term, term clustering

4 Human-computer interface development: concepts and systems for its management

H. Rex Hartson, Deborah Hix

March 1989 **ACM Computing Surveys (CSUR)**, Volume 21 Issue 1

Publisher: ACM Press

Full text available: pdf(7.97 MB) Additional Information: full citation, abstract, references, citings, index terms, review

Human-computer interface management, from a computer science viewpoint, focuses on the process of developing quality human-computer interfaces, including their representation, design, implementation, execution, evaluation, and maintenance. This survey presents important concepts of interface management: dialogue independence, structural modeling, representation, interactive tools, rapid prototyping, development methodologies, and control structures. *Dialogue independence* is th ...

5 Data clustering: a review

A. K. Jain, M. N. Murty, P. J. Flynn

September 1999 **ACM Computing Surveys (CSUR)**, Volume 31 Issue 3

Publisher: ACM Press

Full text available: pdf(636.24 KB) Additional Information: full citation, abstract, references, citings, index terms, review

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic co ...

Keywords: cluster analysis, clustering applications, exploratory data analysis, incremental clustering, similarity indices, unsupervised learning

6 XIRQL: An XML query language based on information retrieval concepts

Norbert Fuhr, Kai Großjohann

April 2004 **ACM Transactions on Information Systems (TOIS)**, Volume 22 Issue 2

Publisher: ACM Press

Full text available: pdf(281.91 KB) Additional Information: full citation, abstract, references, citings, index terms

XIRQL ("circle") is an XML query language that incorporates imprecision and vagueness for both structural and content-oriented query conditions. The corresponding uncertainty is handled by a consistent probabilistic model. The core features of XIRQL are (1) document ranking based on index term weighting, (2) specificity-oriented search for retrieving the most relevant parts of documents, (3) datatypes with vague predicates for dealing with specific types of content and (4) structural vagueness f ...

Keywords: Path algebra, XML, XQuery, probabilistic retrieval, ranked retrieval, vague predicates

7 Multi-answer-focused multi-document summarization using a question-answering engine

Tatsunori Mori, Masanori Nozawa, Yoshiaki Asada
September 2005 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 4 Issue 3
Publisher: ACM Press
Full text available: pdf(635.10 KB) Additional Information: full citation, abstract, references, index terms

In recent years, answer-focused summarization has gained attention as a technology complementary to information retrieval and question answering. In order to realize multi-document summarization focused by multiple questions, we propose a method to calculate sentence importance using scores, for responses to multiple questions, generated by a Question-Answering engine. Further, we describe the integration of this method with a generic multi-document summarization system. The evaluation results d ...

Keywords: Information gain ratio, maximal marginal relevance, question-answering engine

8 Self-indexing inverted files for fast text retrieval

Alistair Moffat, Justin Zobel
October 1996 **ACM Transactions on Information Systems (TOIS)**, Volume 14 Issue 4
Publisher: ACM Press
Full text available: pdf(484.52 KB) Additional Information: full citation, abstract, references, citings, index terms

Query-processing costs on large text databases are dominated by the need to retrieve and scan the inverted list of each query term. Retrieval time for inverted lists can be greatly reduced by the use of compression, but this adds to the CPU time required. Here we show that the CPU component of query response time for conjunctive Boolean queries and for informal ranked queries can be similarly reduced, at little cost in terms of storage, by the inclusion of an internal index in each compress ...

9 Web document clustering: a feasibility demonstration

Oren Zamir, Oren Etzioni
August 1998 **Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(1.43 MB) Additional Information: full citation, references, citings, index terms

10 Document and images analysis: INFTY: an integrated OCR system for mathematical documents

Masakazu Suzuki, Fumikazu Tamari, Ryoji Fukuda, Seiichi Uchida, Toshihiro Kanahori
November 2003 **Proceedings of the 2003 ACM symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(322.41 KB) Additional Information: full citation, abstract, references, index terms

An integrated OCR system for mathematical documents, called INFTY, is presented. INFTY consists of four procedures, i.e., layout analysis, character recognition, structure analysis of mathematical expressions, and manual error correction. In those procedures, several novel techniques are utilized for better recognition performance. Experimental results on about 500 pages of mathematical documents showed high character recognition rates on both mathematical expressions and ordinary texts, and suf ...

Keywords: character and symbol recognition, mathematical OCR, structure analysis of mathematical expressions

11 Papers: Identifying, the coding system and language, of on-line documents on the Internet
Gen-itiro Kikui
August 1996 **Proceedings of the 16th conference on Computational linguistics - Volume 2**
Publisher: Association for Computational Linguistics
Full text available: pdf(523.04 KB) Additional Information: full citation, abstract, references

This paper proposes a new algorithm that simultaneously identifies the coding system and language of a code string fetched from the Internet, especially World-Wide Web. The algorithm uses statistic language models to select the correctly decoded string as well as to determine the language. The proposed algorithm covers 9 languages and 11 coding systems used in Eastern Asia and Western Europe. Experimental results show that the level of accuracy of our algorithm is over 95% for 640 on-line docume ...

12 Selected IR-Related Dissertation Abstracts
September 1991 **ACM SIGIR Forum**, Volume 25 Issue 2
Publisher: ACM Press
Full text available: pdf(2.75 MB) Additional Information: full citation, abstract

The following are citations selected by title and abstract as being related to Information Retrieval (IR), resulting from a computer search, using BRS Information Technologies, of the Dissertation Abstracts Online database produced by University Microfilms International (UMI). Included are UMI order number, title, author, degree, year, institution; number of pages, one or more Dissertation Abstracts International (DAI) subject descriptors chosen by the author, and abstract. Unless otherwise spec ...

13 Document Databases: Requirements for XML document database systems
Airi Salminen, Frank Wm. Tompa
November 2001 **Proceedings of the 2001 ACM Symposium on Document engineering**
Publisher: ACM Press
Full text available: pdf(141.89 KB) Additional Information: full citation, abstract, references, citings, index terms

The shift from SGML to XML has created new demands for managing structured documents. Many XML documents will be transient representations for the purpose of data exchange between different types of applications, but there will also be a need for effective means to manage persistent XML data as a database. In this paper we explore requirements for an XML database management system. The purpose of the paper is not to suggest a single type of system covering all necessary features. Instead the pur ...

Keywords: XML, XML database systems, data definition, data manipulation, data modelling, structured documents

14 Searching in metric spaces
Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, José Luis Marroquín
September 2001 **ACM Computing Surveys (CSUR)**, Volume 33 Issue 3
Publisher: ACM Press
Full text available: pdf(916.04 KB) Additional Information: full citation, abstract, references, citings, index terms

The problem of searching the elements of a set that are close to a given query element under some similarity criterion has a vast number of applications in many branches of computer science, from pattern recognition to textual and multimedia information retrieval. We are interested in the rather general case where the similarity criterion defines a metric space, instead of the more restricted case of a vector space. Many solutions have been proposed in different areas, in many cases without cros ...

Keywords: Curse of dimensionality, nearest neighbors, similarity searching, vector spaces

15 An XML query engine for network-bound data
Zachary G. Ives, A. Y. Halevy, D. S. Weld
December 2002 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 11 Issue 4
Publisher: Springer-Verlag New York, Inc.
Full text available: pdf(351.86 KB) Additional Information: full citation, abstract, citings, index terms

XML has become the lingua franca for data exchange and integration across administrative and enterprise boundaries. Nearly all data providers are adding XML import or export capabilities, and standard XML Schemas and DTDs are being promoted for all types of data sharing. The ubiquity of XML has removed one of the major obstacles to integrating data from widely disparate sources - namely, the heterogeneity of data formats. However, general-purpose integration of data across the wide are a also re ...

Keywords: Data integration, Data streams, Query processing, Web and databases, XML

16 Special issue: AI in engineering
D. Sriram, R. Joobbani
April 1985 **ACM SIGART Bulletin**, Issue 92
Publisher: ACM Press
Full text available: pdf(8.79 MB) Additional Information: full citation, abstract

The papers in this special issue were compiled from responses to the announcement in the July 1984 issue of the SIGART newsletter and notices posted over the ARPAnet. The interest being shown in this area is reflected in the sixty papers received from over six countries. About half the papers were received over the computer network.

17 Vector-based natural language call routing
Jennifer Chu-Carroll, Bob Carpenter
September 1999 **Computational Linguistics**, Volume 25 Issue 3
Publisher: MIT Press
Full text available: pdf(1.87 MB) Publisher Site Additional Information: full citation, abstract, references, citings

This paper describes a domain-independent, automatically trained natural language call router for directing incoming calls in a call center. Our call router directs customer calls based on their response to an open-ended *How may I direct your call?* prompt. Routing behavior is trained from a corpus of transcribed and hand-routed calls and then carried out using vector-based information retrieval techniques. Terms consist of *n*-gram sequences of morphologically reduced content words, ...

18 Inverted files versus signature files for text indexing
Justin Zobel, Alistair Moffat, Kotagiri Ramamohanarao
December 1998 **ACM Transactions on Database Systems (TODS)**, Volume 23 Issue 4
Publisher: ACM Press
Full text available: pdf(243.62 KB) Additional Information: full citation, abstract, references, citings, index terms

Two well-known indexing methods are inverted files and signature files. We have undertaken a detailed comparison of these two approaches in the context of text indexing, paying particular attention to query evaluation speed and space requirements. We have examined their relative performance using both experimentation and a refined approach to modeling of signature files, and demonstrate that inverted files are distinctly superior to signature files. Not only can inverted files be used to ev ...

Keywords: indexing, inverted files, performance, signature files, text databases, text indexing

19 A new character-based indexing method using frequency data for Japanese documents
Ogawa Yasushi, Iwasaki Masajirou
July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(891.01 KB) Additional Information: full citation, references, citings, index terms

20 Fast and quasi-natural language search for gigabytes of Chinese texts
Lee-Feng Chien
July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(820.58 KB) Additional Information: full citation, references, citings, index terms

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next

PORTAL
USPTO

Subscribe (Full Service) Register (Limited Service, Free) Login

Search: ● The ACM Digital Library ○ The Guide

evaluation documents character strings retrieving smilar docur SEARCH

THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

Terms used
evaluation documents character strings retrieving smilar documents

Found **43,086** of **171,143**

Sort results by: relevance
Display results: expanded form

Save results to a Binder
Search Tips
□ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next
Best 200 shown

Relevance scale

1 A new character-based indexing method using frequency data for Japanese documents
Ogawa Yasushi, Iwasaki Masajirou
July 1995 **Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(891.01 KB) Additional Information: full citation, references, citings, index terms

2 Access by content of documents in an office information system
C. Jimenez Guarin
May 1988 **Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(1.47 MB) Additional Information: full citation, abstract, references, index terms

This paper presents the integration of retrieval functions of an Information Retrieval System, IOTA, in an Office Information Server. Besides the linear scanning of the text (using a software and a hardware filter), two access methods are proposed. The first one is based on a simple indexing of documents based on signatures. Here, texts are treated as character strings. We call this method Textual Search. The second one is based on the extention of Signature Methods ...

3 A comparison of Chinese document indexing strategies and retrieval models
Robert W. P. Luk, K. L. Kwok
September 2002 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 1 Issue 3
Publisher: ACM Press
Full text available: pdf(419.42 KB) Additional Information: full citation, abstract, references, index terms

With the advent of the Internet and intranets, substantial interest is being shown in Asian language information retrieval; especially in Chinese, which is a good example of an Asian ideographic language (other examples include Japanese and Korean). Since, in this type of language, spaces do not delimit words, an important issue is which index terms should be extracted from documents. This issue also has wider implications for indexing other languages such as agglutinating languages (e.g., Finni ...

Keywords: Chinese information retrieval, comparison, indexing strategies

4 Evaluation of model-based retrieval effectiveness with OCR text
Kazem Taghva, Julie Borsack, Allen Condit
January 1996 **ACM Transactions on Information Systems (TOIS)**, Volume 14 Issue 1
Publisher: ACM Press
Full text available: pdf(2.02 MB) Additional Information: full citation, abstract, references, citings, index terms, review

We give a comprehensive report on our experiments with retrieval from OCR-generated text using systems based on standard models of retrieval. More specifically, we show that average precision and recall is not affected by OCR errors across systems for several collections. The collections used in these experiments include both actual OCR-generated text and standard information retrieval collections corrupted through the simulation of OCR errors. Both the actual and simulation experiments inc ...

Keywords: error correction, feedback, optical character recognition, ranking algorithms

5 XRel: a path-based approach to storage and retrieval of XML documents using relational databases
August 2001 **ACM Transactions on Internet Technology (TOIT)**, Volume 1 Issue 1
Publisher: ACM Press
Full text available: pdf(264.27 KB) Additional Information: full citation, abstract, references, citings, index terms, review

This article describes XRel, a novel approach for storage and retrieval of XML documents using relational databases. In this approach, an XML document is decomposed into nodes on the basis of its tree structure and stored in relational tables according to the node type, with path information from the root to each node. XRel enables us to store XML documents using a fixed relational schema without any information about DTDs and also to utilize indices such as the B+
Keywords: XML query, XPath, text markup, text tagging

6 A Chinese dictionary construction algorithm for information retrieval
Honglan Jin, Kam-Fai Wong
December 2002 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 1 Issue 4
Publisher: ACM Press
Full text available: pdf(133.47 KB) Additional Information: full citation, abstract, references, index terms

In this article we propose a method for constructing, from raw Chinese text, a statistics-based automatic dictionary. The method makes use of local statistical information (i.e., data within a document) to identify and discard repeated string patterns, which, at an earlier stage, were substrings of legitimate words. Global statistical information (which exists throughout the entire corpus) and contextual constraints are then used for further filtering. The method can be used to alleviate the out ...

Keywords: Chinese information retrieval, automatic word extraction, dictionary construction

7 String Match and Text Extraction: Improved string matching under noisy channel conditions

Kevyn Collins-Thompson, Charles Schweizer, Susan Dumais
October 2001 **Proceedings of the tenth international conference on Information and knowledge management**
Publisher: ACM Press
Full text available: pdf(1.71 MB) Additional Information: full citation, abstract, references, index terms

Many document-based applications, including popular Web browsers, email viewers, and word processors, have a 'Find on this Page' feature that allows a user to find every occurrence of a given string in the document. If the document text being searched is derived from a noisy process such as optical character recognition (OCR), the effectiveness of typical string matching can be greatly reduced. This paper describes an enhanced string-matching algorithm for degraded text that improves recall, whi ...

Keywords: approximate string matching, information retrieval evaluation, noisy channel model, optical character recognition

8 Query processing in a multimedia document system
Elisa Bertino, Fausto Rabbiti, Simon Gibbs
January 1988 **ACM Transactions on Information Systems (TOIS)**, Volume 6 Issue 1
Publisher: ACM Press
Full text available: pdf(2.94 MB) Additional Information: full citation, abstract, references, citings, index terms, review

Query processing in a multimedia document system is described. Multimedia documents are information objects containing formatted data, text, image, graphics, and voice. The query language is based on a conceptual document model that allows the users to formulate queries on both document content and structure. The architecture of the system is outlined, with focus on the storage organization in which both optical and magnetic devices can coexist. Query processing and the different strategies ...

9 Exploiting parallelism in pattern matching: an information retrieval application
Victor Wing-Kit Mak, Kuo Chu Lee, Ophir Frieder
January 1991 **ACM Transactions on Information Systems (TOIS)**, Volume 9 Issue 1
Publisher: ACM Press
Full text available: pdf(1.42 MB) Additional Information: full citation, abstract, references, citings, index terms, review

We propose a document-searching architecture based on high-speed hardware pattern matching to increase the throughput of an information retrieval system. We also propose a new parallel VLSI pattern-matching algorithm called the Data Parallel Pattern Matching (DPPM) algorithm, which serially broadcasts and compares the pattern to a block of data in parallel. The DPPM algorithm utilizes the high degree of integration of VLSI technology to attain very high-speed processing through parallelism. ...

Keywords: DPPM, pattern matcher

10 Character cluster based Thai information retrieval
Thanaruk Theeramunkong, Virach Sornlertlamvanich, Thanasan Tanhermhong, Wirat Chinnan
November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**
Publisher: ACM Press
Full text available: pdf(516.94 KB) Additional Information: full citation, abstract, references

Some languages including Thai, Japanese and Chinese do not have explicit word boundary. This causes the problem of word boundary ambiguity that results in decreasing the accuracy of information retrieval. This paper proposes a new technique so-called character clustering to reduce the ambiguity of word boundary in Thai documents and hence improve searching efficiency. To investigate the efficiency, a set of experiments using Thai newspapers is conducted in both non-indexing and indexing searc ...

Keywords: Thai document, character cluster, indexing and non-indexing information retrieval

11 Document engineering (DE): Performance evaluation for text processing of noisy inputs
Daniel Lopresti
March 2005 **Proceedings of the 2005 ACM symposium on Applied computing SAC '05**
Publisher: ACM Press
Full text available: pdf(110.60 KB) Additional Information: full citation, abstract, references, index terms

We investigate the problem of evaluating the performance of text processing algorithms on inputs that contain errors as a result of optical character recognition. A new hierarchical paradigm is proposed based on approximate string matching, allowing each stage in the processing pipeline to be tested, the error effects analyzed, and possible solutions suggested.

Keywords: optical character recognition, part-of-speech tagging, performance evaluation, sentence boundary detection, tokenization

12 An algorithm for retrieving indexed documents and its application
Allan J. Humphrey, Shelby L. Brumelle
January 1966 **Proceedings of the 1966 21st national conference**
Publisher: ACM Press
Full text available: pdf(334.66 KB) Additional Information: full citation, abstract, index terms

In recent years the design and development of computerized information storage and retrieval systems has received widespread attention. Such systems have been applied to a broad spectrum of commercial, industrial, governmental, and military activities. One area where such systems have been particularly valuable is that of literature document retrieval. In a literature retrieval system the computer output generally consists of such information as the title, authors, accession number, bibliog ...

13 Document Formatting Systems: Survey, Concepts, and Issues
Richard Furuta, Jeffrey Scofield, Alan Shaw
September 1982 **ACM Computing Surveys (CSUR)**, Volume 14 Issue 3
Publisher: ACM Press
Full text available: pdf(5.36 MB) Additional Information: full citation, references, citings, index terms

14 Fast detection of communication patterns in distributed executions
Thomas Kunz, Michiel F. H. Seuren
November 1997 **Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research**
Publisher: IBM Press
Full text available: pdf(4.21 MB) Additional Information: full citation, abstract, references, index terms

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

15 A browsing tool of multi-lingual documents for users without multi-lingual fonts
Tetsuo Sakaguchi, Akira Maeda, Takehisa Fujita, Shigeo Sugimoto, Koichi Tabata
April 1996 **Proceedings of the first ACM international conference on Digital libraries**
Publisher: ACM Press
Full text available: pdf(789.68 KB) Additional Information: full citation, references, citings, index terms

16 On Chinese text retrieval
Jian-Yun Nie, Martin Brisebois, Xiaobo Ren
August 1996 **Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(925.66 KB) Additional Information: full citation, references, citings, index terms

17 Experiments on incorporating syntactic processing of user queries into a document retrieval strategy
A. F. Smeaton, C. J. van Rijsbergen
May 1988 **Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval**
Publisher: ACM Press
Full text available: pdf(1.94 MB) Additional Information: full citation, abstract, references, citings, index terms

Traditional information has relied on the extensive use of statistical parameters in the implementation of retrieval strategies. This paper sets out to investigate whether linguistic processes can be used as part of a document retrieval strategy. This is done by predefining a level of syntactic analysis of user queries only, to be used as part of the retrieval process. A large series of experiments on an experimental test collection are reported which use a parser for noun phrases as part o ...

18 Proximal nodes: a model to query document databases by content and structure
Gonzalo Navarro, Ricardo Baeza-Yates
October 1997 **ACM Transactions on Information Systems (TOIS)**, Volume 15 Issue 4
Publisher: ACM Press
Full text available: pdf(550.43 KB) Additional Information: full citation, abstract, references, citings, index terms, review

A model to query document databases by both their content and structure is presented. The goal is to obtain a query language that is expressive in practice while being efficiently implementable, features not present at the same time in previous work. The key ideas of the model are a set-oriented query language based on operations on nearby structure elements of one or more hierarchies, together with content and structural indexing and bottom-up evaluation. The model is evaluated in regard t ...

Keywords: expressivity and efficiency of query languages, hierarchical documents, structured text, text algebras

19 Content-based language models for spoken document retrieval

Hsin-min Wang, Berlin Chen

November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**

Publisher: ACM Press

Full text available: pdf(662.94 KB) Additional Information: full citation, abstract, references

Spoken document retrieval (SDR) has been extensively studied in recent years because of its potential use in navigating large multimedia collections in the near future. This paper presents a novel concept of applying the content-based language models to spoken document retrieval. In an example task for retrieval of Mandarin broadcast news, the content-based language models either trained with the automatic transcriptions of the spoken documents or adapted from the baseline language models usi ...

Keywords: content-based language models, speech recognition, spoken document retrieval

20 A backend machine architecture for information retrieval

Amar Mukhopadhyay

June 1980 **Proceedings of the 3rd annual ACM conference on Research and development in information retrieval**

Publisher: Butterworth & Co.

Full text available: pdf(525.26 KB) Additional Information: full citation, references

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next

Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player

WEST Search History

Hide Items | Restore | Clear | Cancel

DATE: Tuesday, March 07, 2006

Hide?	Set Name	Query	Hit Count
		DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=ADJ	
□	L35	L31 and (fitness same unfitness)	1
□	L34	L23 and (fitness same unfitness)	0
□	L33	L32 and ((similar near5 document$1) same (calculat$3 near5 degree))	5
□	L32	L31 and (evaluat$3 near5 document$1)	14
□	L31	seed document$1	68
□	L30	'seed document' and (similar$5 near5 document$1) and (search$3 near5 document$1) and (search$3 near5 engine$1) and @py<=2003	7
□	L29	L28 and (vector near5 space)	6
□	L28	L27 and vector$1	8
□	L27	L26 and weight$1	10
□	L26	L25 and (character near5 string$1)	10
□	L25	L24 and (calculat$3 near5 similar$5)	18
□	L24	L23 and (boolean and evaluat$3 and algorithm$1)	39
□	L23	(compar$3 near5 document$1) and (retriev$3 near5 document$1) and (document$1 near5 frequenc$3) and threshold and @py<=2003	186
□	L22	(compar$3 near5 document$1) and (rectriev$3 near5 document$1) and (document$1 near5 frequenc$3) and threshold and @py<=2003	0
□	L21	(document$3 near5 similarities) and (search$3 near5 engine$1) and (user$1 near5 profil$3) and (compar$3 near5 document$1) and (frequency and weight$1 and vector$1 and (character near5 string$1) and boolean and retriev$3 and stor$3 and (evaluat$3 near5 document$1)) and @py<=2002	2
□	L20	L19 and (document$1 near5 weight$1)	7
□	L19	L18 and (document$1 or frequen$4)	19
□	L18	L16 and ((document$1 or topic$1) near5 (similar$5))	19
□	L17	L16 and ((document$1 or topic$1) near5 (smillar$5))	0
□	L16	(search$3 and engine$1 and user$1 and topic$1 and document$1 and threshold and profile$1 and degree and calculat$3 and vector$1 and weight$1) and @py<=2002	35
□	L15	L11 and (degree same document$1)	5
□	L14	L11 and (degree same calculat$3)	0
□	L13	L11 and (degree same fitness)	0
□	L12	L11 and (degree near5 fitness)	0

□	L11	L10 and weight$1	18
□	L10	L9 and (document near5 similar$5)	21
□	L9	(search$3 and engine$1 and document$1 and boolean and vector$1 and string$1 and character$1 and retriev$3 and exclud$3 and user$1) and @py<=2002	77
□	L8	(document$1 and character$1 and string$1 and (filter$1 or search$3 or query$3) and engine$1 and (similar near5 document$1) and evaluat$3 and (exclud$3 near5 document$1) and weight$1 and boolean) and @py<=2002	5
□	L7	6415282.uref.	14
□	L6	L5 and n$gram$1	1
□	L5	L4 and (document$1 near5 weight$1)	5
□	L4	L2 and boolean	20
□	L3	L2 and (evaluat$3 near5 document$1)	2
□	L2	(document$1 and seed and calculat$3 and similarit$3 and vector and space and search$3 and engine$1) and @py<=2002	39
□	L1	(search$3 and engine$1 and user$1 and boolean and document$1 and character and strings and evaluat$3 and calculat$3 and similarit$3 and vector and space and filter$3 and inconsistenc$3) and @py<=2002	2

END OF SEARCH HISTORY